Pattern Recognition Letters 31 (2010) 2489–2492

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Improved direct LDA and its application to DNA microarray gene expression data

Kuldip K. Paliwal [a], Alok Sharma [b,c,*]

[a] School of Engineering, Griffith University, QLD-4111, Australia
[b] Signal Processing Lab, Griffith University, QLD-4111, Australia
[c] School of Engineering and Physics, University of the South Pacific, Suva, Fiji

ARTICLE INFO

Article history:
Received 12 January 2010
Available online 21 August 2010
Communicated by T. Vasilakos

Keywords:
Direct linear discriminant analysis
Small sample size problem
Dimensionality reduction

ABSTRACT

The direct linear discriminant analysis (DLDA) technique is a well known technique for dimensionality reduction. It can overcome the small sample size problem. However, its performance is limited. In this paper we address its drawbacks and propose an improvement of the DLDA technique. The experiment is conducted on several DNA microarray gene expression datasets and the performance (in terms of classification accuracy) of the proposed improvement of the technique is reported at 91.1% which is very promising.

1. Introduction

Direct linear discriminant analysis (DLDA) is an important dimensionality reduction technique for solving small sample size problem (Yu and Yang, 2001). In the DLDA method, the dimensionality is reduced in two stages. In the first stage, a transformation matrix is computed to transform the training samples to the range space of $\mathbf{S}_B$. In the second stage, the dimensionality of this transformed samples is further transformed by some regulating matrices. The benefit of DLDA technique is that it does not require PCA transformations to reduce the dimensionality as required by other techniques like Fisherface (or PCA + LDA) technique (Swets and Weng, 1996; Belhumeur et al., 1997). Though several extensions of DLDA technique (Song et al., 2007; Lu et al., 2003; Lu et al., 2005) have been reported in the literature, there still exist some important aspects in regards to its performance.

In the first stage of the DLDA technique the transformation matrix is computed by weighting the eigenvectors of the range space of $\mathbf{S}_B$ with the inverse eigenvalues of the range space of $\mathbf{S}_B$. Theoretically, there are two main drawbacks of this procedure: (1) the null space of $\mathbf{S}_B$ is ignored in this stage; and (2) since the inverse of the eigenvalues of $\mathbf{S}_B$ are utilized in the computation operation, the eigenvectors corresponding to larger eigenvalues of $\mathbf{S}_B$ are de-emphasized which are more useful in discriminant analysis. In the second stage the dimensionality is reduced to some value h by choosing h eigenvectors of the within-class scatter matrix $\hat{\mathbf{S}}_W$ (of the transformed training samples) corresponding to the lowest eigenvalues. This stage also requires inverse computation of the eigenvalues of $\hat{\mathbf{S}}_W$ which has a drawback that zero eigenvalues of $\hat{\mathbf{S}}_W$ may exist which could cause problem in finding its inverse.

In this paper, we propose an improved DLDA technique which addresses these drawbacks of DLDA technique. In the improved DLDA technique we first decompose $\mathbf{S}_W$ matrix into its eigenvalues and eigenvectors instead of $\mathbf{S}_B$ matrix as of DLDA technique. Here we utilize both its null space and range space information by approximating $\mathbf{S}_W^{-1}$ by a well deterministic substitution. Then we diagonalize $\mathbf{S}_B$ matrix using regulating matrices. The proposed improvement has been experimented on several DNA microarray gene expression datasets and significant advancement (in terms of classification accuracy) over DLDA technique has been observed.

2. The DLDA technique

In order to describe the DLDA technique, we first define some notations. Let $\bar{\chi}$ be a set of n training vectors (samples or patterns) in a d-dimensional feature space, $\Omega = \{\omega_i: i = 1, 2, \ldots, c\}$ be the finite set of c states of nature or class labels where ω_i denotes the ith class label. The set $\bar{\chi}$ can be subdivided into c subsets $\bar{\chi}_1, \bar{\chi}_2, \ldots, \bar{\chi}_c$ where each subset $\bar{\chi}_i$ belongs to ω_i and consists of n_i number of samples such that:

$$n = \sum_{i=1}^{c} n_i$$

The training samples or patterns of set $\bar{\chi}$ can be written as:

$$\bar{\chi} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\} \quad \text{where } \mathbf{x}_j \in \mathbf{R}^d$$
$$\bar{\chi}_i \subset \bar{\chi} \text{ and } \bar{\chi}_1 \cup \bar{\chi}_2 \cup \ldots \cup \bar{\chi}_c = \bar{\chi}$$

* Corresponding author at: School of Engineering and Physics, University of the South Pacific, Suva, Fiji. Tel.: +679 323 2870; fax: +679 323 1538. Tel.: +61 7 3735 6536 (Australia).
E-mail address: sharma_al@usp.ac.fj (A. Sharma).

0167-8655/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2010.08.003

Let $\boldsymbol{\mu}_j$ be the centroid of $\bar{\chi}_i$ and $\boldsymbol{\mu}$ be the centroid of $\bar{\chi}$, then the between class scatter matrix is given as:

$$\mathbf{S}_B = \sum_{j=1}^{c} n_j(\boldsymbol{\mu}_j - \boldsymbol{\mu})(\boldsymbol{\mu}_j - \boldsymbol{\mu})^{\mathrm{T}} \tag{1}$$

It can be observed from Eq. (1) that $\mathbf{S}_B$ is the sum of c matrices of rank one or less, and because only $c-1$ of these are independent, the rank of $\mathbf{S}_B$ is $b = rank(\mathbf{S}_B) \leqslant c-1$ (Duda and Hart, 1973).

The within-class scatter matrix which is the sum of c scatter matrices can be defined by:

$$\mathbf{S}_w = \sum_{i=1}^{c} \mathbf{S}_i$$

where $\mathbf{S}_i = \sum_{x \in \bar{\chi}_i} (\mathbf{x} - \boldsymbol{\mu}_i)(\mathbf{x} - \boldsymbol{\mu}_i)^{\mathrm{T}}$

The rank of $\mathbf{S}_w$ (for $n \geqslant c$) is:

$$rank(\mathbf{S}_w) \leqslant n - c$$

Let $\mathbf{S}_B = \mathbf{H}_B \mathbf{H}_B^{\mathrm{T}}$ then from SVD of $\mathbf{H}_B$ we get:

$$\mathbf{U}^{\mathrm{T}} \mathbf{S}_B \mathbf{U} = \mathbf{D}_B^2 \tag{2}$$

where $\mathbf{D}_B = \begin{bmatrix} \boldsymbol{\Lambda}_B & 0 \\ 0 & 0 \end{bmatrix}$ and $\boldsymbol{\Lambda}_B \in \mathbf{R}b^{\times b}$

Let $\mathbf{U} = [\mathbf{U}_r, \mathbf{U}_n]$ where eigenvectors $\mathbf{U}_r \in \mathbf{R}^{d \times b}$ corresponds to the range space of $\mathbf{S}_B$ (i.e., $\boldsymbol{\Lambda}_B$) and eigenvectors $\mathbf{U}_n \in \mathbf{R}^{d \times (d-b)}$ corresponds to the null space of $\mathbf{S}_B$.

In the first stage of DLDA technique, the transformation matrix is computed as follows. The eigenvectors corresponding to the null space of $\mathbf{S}_B$ are discarded which would modify Eq. (2) as:

$$\mathbf{U}_r^{\mathrm{T}} \mathbf{S}_B \mathbf{U}_r = \boldsymbol{\Lambda}_B^2 \tag{3}$$

$$\text{or} \quad \boldsymbol{\Lambda}_B^{-1} \mathbf{U}_r^{\mathrm{T}} \mathbf{S}_B \mathbf{U}_r \boldsymbol{\Lambda}_B^{-1} = \mathbf{I}_{b \times b} \tag{4}$$

The transformation $\mathbf{U}_r \boldsymbol{\Lambda}_B^{-1}$ is used to transform $\mathbf{S}_w$ matrix as:

$$\hat{\mathbf{S}}_w = \boldsymbol{\Lambda}_B^{-1} \mathbf{U}_r^{\mathrm{T}} \mathbf{S}_w \mathbf{U}_r \boldsymbol{\Lambda}_B^{-1} \tag{5}$$

In the second stage of the technique the transforms are regulated as follows. Let the eigenvalue decomposition of $\hat{\mathbf{S}}_w$ be $\mathbf{F} \boldsymbol{\Sigma}_w^2 \mathbf{F}^{\mathrm{T}}$, then we can write $\boldsymbol{\Sigma}_w^2$ as:

$$\mathbf{F}^{\mathrm{T}} \hat{\mathbf{S}}_w \mathbf{F} = \boldsymbol{\Sigma}_w^2$$

$$\text{or} \quad \boldsymbol{\Sigma}_w^{-1} \mathbf{F}^{\mathrm{T}} \hat{\mathbf{S}}_w \mathbf{F} \boldsymbol{\Sigma}_w^{-1} = \mathbf{I}_{b \times b} \tag{6}$$

The value of $\hat{\mathbf{S}}_w$ (from Eq. (5)) can be substituted in Eq. (6), we will get:

$$\boldsymbol{\Sigma}_w^{-1} \mathbf{F}^{\mathrm{T}} \boldsymbol{\Lambda}_B^{-1} \mathbf{U}_r^{\mathrm{T}} \mathbf{S}_w \mathbf{U}_r \boldsymbol{\Lambda}_B^{-1} \mathbf{F} \boldsymbol{\Sigma}_w^{-1} = \mathbf{I}_{b \times b}$$

Therefore the orientation matrix $\mathbf{W} \in \mathbf{R}^{d \times b}$ of DLDA technique can be given as:

$$\mathbf{W} = \mathbf{U}_r \boldsymbol{\Lambda}_B^{-1} \mathbf{F} \boldsymbol{\Sigma}_w^{-1} \tag{7}$$

3. Drawbacks of DLDA technique

In this section we discuss the drawbacks of DLDA technique. It can be seen from Eq. (3) that only the range space of $\mathbf{S}_B$ is retained in the computation process and its null space has been discarded. The null space of $\mathbf{S}_B$ could also have crucial information useful for classification (see Appendix A for experimental details). Discarding the null space of $\mathbf{S}_B$ would reduce the inherited information and subsequently the discriminant power would reduce.

Next drawback is the inverse operation of the eigenvalues of $\mathbf{S}_B$ in Eq. (4). The multiplication of $\boldsymbol{\Lambda}_B^{-1}$ with $\mathbf{U}_r$ would deemphasize the eigenvectors corresponding to larger eigenvalues of $\mathbf{S}_B$ (Song et al., 2007). This operation may adversely affect the classification performance.

Furthermore, the next drawback is due to the inverse operation of $\boldsymbol{\Sigma}_w$ in Eq. (6) as it cannot be guaranteed that all the eigenvalues of $\hat{\mathbf{S}}_w$ are non zero. The zero values in $\boldsymbol{\Sigma}_w$ may cause problem in finding its inverse.

4. Improved DLDA technique

The improved DLDA technique can be described in two stages. In the first stage, the $\mathbf{S}_w$ matrix is decomposed into its eigenvalues and eigenvectors instead of $\mathbf{S}_B$ matrix as of DLDA technique. Here the range space of $\mathbf{S}_w$ and the null space of $\mathbf{S}_w$ are used for further processing. In order to use both the spaces, we approximate the inverse of the square root of eigenvalues of $\mathbf{S}_w$ with a well deterministic substitution. In the inverse computation of the eigenvalues of $\mathbf{S}_w$ the small eigenvalues become large and the large eigenvalues become small. To get this inverse realization, we substituted the inverse operation by the subtraction operation as follows. The square root of eigenvalues of $\mathbf{S}_w$ is subtracted from $\alpha \mathbf{I}$, where α is the largest square root of eigenvalue of $\mathbf{S}_w$ and $\mathbf{I}$ is an identity matrix of the size of $\mathbf{S}_w$. This will give a matrix $\mathbf{D}_\alpha$. The matrix $\mathbf{D}_\alpha$ is the substitution of the inverse of square root of eigenvalues of $\mathbf{S}_w$. This operation has an advantage over the DLDA technique (where the null space of $\mathbf{S}_B$ has been discarded) as the null space of $\mathbf{S}_B$ is not discarded therefore some information crucial for discriminant analysis will be retained. Secondly, $\boldsymbol{\Lambda}_B^{-1}$ is not used to weight eigenvectors of $\mathbf{S}_B$ which avoids the instance of deemphasizing the eigenvectors corresponding to larger eigenvalues of $\mathbf{S}_B$. In the second stage, the $\mathbf{S}_B$ matrix is diagonalized using the regulating matrices which would give the orientation matrix (of the improved DLDA technique) $\mathbf{W}$. This process does not require the inverse computation of diagonal matrix $\boldsymbol{\Sigma}_w$, therefore zero values of diagonal matrix would not cause any problem. The mathematical description of the improved DLDA technique is given as follows.

The matrix $\mathbf{S}_w$ is decomposed as:

$$\mathbf{S}_w = \mathbf{U}_w \mathbf{D}_w^2 \mathbf{U}_w^{\mathrm{T}}, \quad \text{where } \mathbf{U}_w \in \mathbf{R}^{d \times d} \quad \text{and } \mathbf{D}_w \in \mathbf{R}^{d \times d}$$

which will give

$$\mathbf{U}_w^{\mathrm{T}} \mathbf{S}_w \mathbf{U}_w = \mathbf{D}_w^2 \tag{8}$$

where $\mathbf{D}_w = \begin{bmatrix} \boldsymbol{\Lambda}_w & 0 \\ 0 & 0 \end{bmatrix}$ and $\boldsymbol{\Lambda}_w \in \mathbf{R}^{w \times w}$ $(w = rank(\mathbf{S}_w))$

If $\mathbf{S}_w$ is a full rank matrix (i.e., $w = d$) then Eq. (8) can be further decomposed as:

$$\mathbf{D}_w^{-1} \mathbf{U}_w^{\mathrm{T}} \mathbf{S}_w \mathbf{U}_w \mathbf{D}_w^{-1} = \mathbf{I}_{d \times d} \tag{9}$$

But since $w < d$, $\mathbf{D}_w^{-1}$ is not possible. So we want to approximate the inverse of $\mathbf{D}_w$ with a matrix $\mathbf{D}_\alpha$. In order to investigate $\mathbf{D}_\alpha$, we represent the square root of eigenvalues of $\mathbf{S}_w$ (i.e., diagonal elements of $\mathbf{D}_w$) as $[\sigma_1, \sigma_2, \ldots, \sigma_w, \sigma_{w+1}, \ldots, \sigma_d]$ where $\sigma_j = 0$ for $j = w+1, \ldots, d$. If the square root of eigenvalues are arranged in descending order then $\sigma_1 > \sigma_2 > \cdots > \sigma_w > \sigma_{w+1} = \cdots = \sigma_d$. The square root of eigenvalues of the inverse of $\mathbf{S}_w$ can be given as:

$$1/\sigma_1 < 1/\sigma_2 < \ldots < 1/\sigma_w < 1/\sigma_{w+1} = \ldots = 1/\sigma_d \tag{10}$$

where $1/\sigma_j \to \infty$ for $j = w+1, \ldots, d$ which are undesirable terms. To get the inverse realization of $\mathbf{D}_w$ we need to do two things in Eq. (10), (1) maintain the order qualitatively, and (2) replace undesirable terms. To satisfy both the requirements we execute subtraction operation as follows:

$$\beta - \sigma_1 < \beta - \sigma_2 < \ldots < \beta - \sigma_w < \beta - \sigma_{w+1} = \ldots = \beta - \sigma_d \tag{11}$$

where $\beta = p\alpha$ and p is any positive quantity. Substituting $q_j = \beta - \sigma_j$ in Eq. (11), we get:

$q_1 < q_2 < \ldots < q_w < q_{w+1} = \ldots = q_d$

Since $\mathbf{S}_W$ is positive semi-definite matrix, the values of $q_j \geqslant 0$ for $j = 1, \ldots, d$. To estimate the value of β, we need to select the value of p. If the value of $p < 1$ then it will give at least one q_j for which $q_j < 0$. This will violate the properties of positive semi-definite matrix $\mathbf{S}_W$. The other option is to take $p \geqslant 1$. In this work we have selected the value of p to be 1 and verified our selection empirically. The performance on this selection obtained was quite promising when compared with DLDA and other techniques. One can, however, select a value of $p > 1$ using cross-validation on training samples which might improve the performance even further. In our case it seems sufficient to use $p = 1$ as it satisfies both the requirements for the inverse realization of S_w matrix. This approximation will give $\mathbf{D}_\alpha = \alpha \mathbf{I}_{d\times d} - \mathbf{D}_w$, where $\alpha = \mathbf{max}(diag(\mathbf{D}_w))$. The lower eigenvalues are emphasized in the inverse operation. If the diagonal elements of $\mathbf{D}_\alpha$ and $\mathbf{D}_w$ are sorted in ascending order (or in descending order) then the trend of the values of diagonal elements of $\mathbf{D}_\alpha$ would be similar to $\mathbf{\Lambda}_w^{-1}$ and the zero eigenvalues of $\mathbf{D}_w$ will be a positive constant in $\mathbf{D}_\alpha$.

Furthermore, the lower eigenvalues of $\mathbf{D}_w$ are not very stable or noisy and in the inverse operation these eigenvalues get emphasized, augmenting the error due to noise. However, the matrix $\mathbf{D}_\alpha$ is obtained from the subtraction operation and hence it has less chances of emphasizing noisy eigenvalues to a large extent. Therefore, $\mathbf{D}_\alpha$ can be assumed to be a robust approximation of $\mathbf{D}_w^{-1}$.

Substituting $\mathbf{D}_w^{-1}$ by $\mathbf{D}_\alpha$ in Eq. (9) we can:

$$\mathbf{D}_\alpha \mathbf{U}_w^{\mathrm{T}} \mathbf{S}_w \mathbf{U}_w \mathbf{D}_\alpha \approx \mathbf{I}_{d\times d}$$

We can now use this $\mathbf{D}_\alpha$ to diagonalize $\mathbf{S}_B$ as:

$$\mathbf{D}_\alpha \mathbf{U}_w^{\mathrm{T}} \mathbf{S}_B \mathbf{U}_w \mathbf{D}_\alpha = \mathbf{F}\mathbf{\Sigma}^2\mathbf{F}^{\mathrm{T}}, \quad \text{where } \mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Lambda}_b & 0 \\ 0 & 0 \end{bmatrix}, \mathbf{F} \in \mathbf{R}^{d\times d}$$

and $\mathbf{\Lambda}_b \in \mathbf{R}^{b\times b}$

Let $\mathbf{F} = [\mathbf{F}_r, \mathbf{F}_n]$ where $\mathbf{F}_r$ corresponds to the range space of $\mathbf{S}_B$ and $\mathbf{F}_n$ corresponds to the null space of $\mathbf{S}_B$. Then:

$$\mathbf{F}_r^{\mathrm{T}} \mathbf{D}_\alpha \mathbf{U}_w^{\mathrm{T}} \mathbf{S}_B \mathbf{U}_w \mathbf{D}_\alpha \mathbf{F}_r = \mathbf{\Lambda}_b^2 \tag{12}$$

It can be seen from Eq. (12) that the required transform can be given as:

$$\mathbf{W} = \mathbf{U}_w \mathbf{D}_\alpha \mathbf{F}_r \tag{13}$$

The orientation matrix $\mathbf{W}$ of Eq. (13) can be viewed as the improved orientation of Eq. (7). Theoretically, the orientation of Eq. (13) would inherit more information crucial for discriminant purpose. However, computationally it could be expensive to derive $\mathbf{W}$ due to the large size of $\mathbf{S}_w$. This can be resolved using a preprocessing step where the feature vectors are first projected onto the lower dimensional space using the range space of total-scatter matrix $\mathbf{S}_T$ (Huang et al., 2002).

The implementation procedure of the improved DLDA technique is summarized in Table 1.

Table 1
Implementation of improved DLDA technique.

Step 1.	Decompose *Sw* into its eigenvalues and eigenvectors, i.e., $\mathbf{S}_w = \mathbf{U}_w \mathbf{D}_w^2 \mathbf{U}_w^{\mathrm{T}}$
Step 2.	Find $\alpha = \max(diag(\mathbf{D}_w))$ and hence compute $\mathbf{D}_\alpha = \alpha \mathbf{I}_{d\times d} - \mathbf{D}_w$
Step 3.	Use $\mathbf{D}_\alpha$ and $\mathbf{U}_w$ to diagonalize $\mathbf{S}_B$, i.e., $\mathbf{D}_\alpha \mathbf{U}_w^{\mathrm{T}} \mathbf{S}_B \mathbf{U}_w \mathbf{D}_\alpha = \mathbf{F}\mathbf{\Sigma}^2\mathbf{F}^{\mathrm{T}}$, where $\mathbf{F} = [\mathbf{F}_r, \mathbf{F}_n]$ such that F_r corresponds to the range space of $\mathbf{S}_B$ and F_n corresponds to the null space of $\mathbf{S}_B$
Step 4.	Compute orientation matrix $\mathbf{W} = \mathbf{U}_w \mathbf{D}_\alpha \mathbf{F}_r$ to project samples into the reduced dimensional feature space

5. Experimental results

The improved DLDA technique is experimented on DNA microarray gene expression data. Eight commonly available datasets are used. In order to compare the performance (in terms of classification accuracy), other techniques (based on similar concepts) like DLDA, PCA + LDA and OLDA (Ye, 2005) are utilized. For DLDA and Improved DLDA a preprocessing step is conducted where the feature space is first transformed to the range space of $\mathbf{S}_T$. In all the cases, the dimension is reduced to $c - 1$ (where c is the number of classes) and the nearest neighbour classifier is used to classify test feature vectors. The description of all the datasets used in the experimentation is given in Table 2.

The techniques are experimented on all the above mentioned datasets and their classification accuracies are depicted in Table 3. There are five columns in Table 3. The first column depicts the datasets used and the subsequent columns depict the classification accuracies in percentage for the techniques used for the experimentation. The average classification accuracies by DLDA technique, PCA + LDA technique and OLDA technique are 79.6%, 84.0% and 86.8%, respectively. The average classification accuracy by improved DLDA technique is noted to be 91.1%. This experimentation shows the improved DLDA technique is outperforming other techniques.

The proposed algorithm is meant for small sample size (SSS) problem and in this work it is demonstrated on DNA microarray gene expression datasets. However, the researchers working on other areas (where SSS problem does exist, e.g., face recognition, gesture recognition, etc.) can be benefited by this algorithm as well.

Table 2
DNA microarray gene expression datasets used in the experimentation.

Datasets	Class	Dimension	Number of training samples	Number of testing samples
Acute leukemia (Golub et al., 1999)	2	7129	38	34
ALL subtype (Yeoh et al., 2002)	7	12,558	215	112
Breast cancer (van't Veer, 2002)	2	24,481	78	19
Lung adenocarcinoma (Beer et al., 2002)	3	7129	44	23
Lung cancer (Gordon et al., 2002)	2	12,533	32	149
MLL leukemia (Armstrong et al., 2002)	3	12,582	57	15
Prostate tumor (Singh et al., 2002)	2	12,600	102	34
SRBCT (Khan et al., 2001)	4	2308	63	20

Table 3
Classification accuracy (in percentage) on the DNA microarray gene expression datasets.

Database	DLDA	OLDA	PCA + LDA technique	Improved DLDA
Acute leukemia	97.1	97.1	100	100.0
ALL subtype	93.8	86.6	80.4	93.8
Breast cancer	42.1	57.9	73.7	73.7
Lung adenocarcinoma	78.1	81.3	81.3	87.5
Lung cancer	72.5	98.0	63.1	97.3
MLL leukemia	100.0	100.0	100	100.0
Prostate tumor	73.5	73.5	73.5	76.5
SRBCT	80.0	100.0	100	100.00
Average	79.6	86.8	84.0	91.1

6. Conclusion

In this paper the improved DLDA technique is proposed. The technique addressed shortfalls of DLDA technique and it has been observed that the proposed technique outperformed DLDA and other similar techniques like PCA + LDA technique and OLDA technique. The experimentation was conducted on eight DNA microarray gene expression datasets and the improved DLDA technique produced average classification accuracy of 91.1% which was much better than the average classification accuracy of the DLDA technique (79.6%).

Appendix A

In this appendix we explicitly show that the null space of $\mathbf{S}_B$ contain some information useful for classification. Here we demonstrate it experimentally for the cancer classification problem. In order to illustrate this, classification accuracy using the null space of $\mathbf{S}_B$ has been given in Table A1. In these experiments, first the data samples have been transformed to the lower dimensional space by using the range space of $\mathbf{S}_T$ as a preprocessing step. Then the null space of $\mathbf{S}_B$ has been utilized to do dimensionality reduction. The nearest neighbour classifier has been used to classify the test feature vectors.

It can be seen from the above table that though the null space of $\mathbf{S}_B$ is less effective, it contain some useful information for cancer classification.

Table A1. Classification accuracy using the null space of $\mathbf{S}_B$.

Database	Classification accuracy in percentage using the null space of $\mathbf{S}_B$
Acute leukemia	76.5
ALL subtype	42.9
Breast cancer	63.2
Lung cancer	97.3
MLL leukemia	33.3
SRBCT	40.0
Average	58.9

References

Armstrong, S.A., Staunton, J.E., Silverman, L.B., Pieters, R., den Boer, M.L., Minden, M.D., Sallan, S.E., Lander, E.S., Golub, T.R., Korsemeyer, S.J., 2002. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nat. Genet. 30, 41–47. [Data Source1: <http://sdmc.lit.org.sg/GEDatasets/Datasets.html>] [Data Source2: <http://www.broad.mit.edu/cgi-bin/cancer/publications/pub_paper.cgi?mode=view&paper_id=63>].

Beer, D.G., Kardia, S.L.R., Huang, C.-C., Giordano, T.J., Levin, A.M., Misek, D.E., Lin, L., Chen, G., Gharib, T.G., Thomas, D.G., Lizyness, M.L., Kuick, R., Hayasaka, S., Taylor, J.M.G., Iannettoni, M.D., Orringer, M.B., Hanash, S., 2002. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat. Med. 8, 816–824. [Data Source: <http://www.dot.ped.med.umich.edu:2000/ourimage/pub/Lung/index.html>].

Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J., 1997. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Trans. Pattern Anal. Machine Intell. 19 (7), 711–720.

Duda, R.O., Hart, P.E., 1973. Pattern Classification and Scene Analysis. Wiley, New York.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D., Lander, E.S., 1999. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. Science 286, 531–537. [Data Source: <http://www.broad.mit.edu/cgi-bin/cancer/publications/pub_paper.cgi?mode=view&paper_id=43>].

Gordon, G.J., Jensen, R.V., Hsiao, L.-L., Gullans, S.R., Blumenstock, J.E., Ramaswamy, S., Richards, W.G., Sugarbaker, D.J., Bueno, R., 2002. Translation of Microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma. Cancer Res. 62, 4963–4967. [Data Source: <http://www.chestsurg.org>].

Huang, R., Liu, Q., Lu, H., Ma, S., 2002. Solving the small sample size problem of LDA. Proc. ICPR 2002 3, 29–32.

Khan, J., Wei, J.S., Ringner, M., Saal, L.H., Ladanyi, M., Westermann, F., Berthold, F., Schwab, M., Antonescu, C.R., Peterson, C., Meltzer, P.S., 2001. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural network. Nat. Med. 7, 673–679. [Data Source: <http://research.nhgri.nih.gov/microarray/Supplement/>].

Lu, J., Plataniotis, K., Venetsanopoulos, A., 2003. Face recognition using kernel direct discriminant analysis algorithms. IEEE Trans. Neural Networks 14 (1), 117–126.

Lu, J., Plataniotis, K., Venetsanopoulos, A., 2005. Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition. Pattern Recognition Lett. 26 (2), 181–191.

Singh, D., Febbo, P.G., Ross, K., Jackson, D.G., Manola, J., Ladd, C., Tamayo, P., Renshaw, A.A., D'Amico, A.V., Richie, J.P., Lander, E.S., Loda, M., Kantoff, P.W., Golub, T.R., Sellers, W.R., 2002. Gene expression correlates of clinical prostate cancer behavior. Cancer Cell 1, 203–209. [Data Source: <http://www.sdmc.lit.org.sg/GEDatasets/Datasets.html#Prostate>].

Song, F., Zhang, D., Wang, J., Liu, H., Tao, Q., 2007. A parameterized direct LDA and its application to face recognition. Neurocomputing 71, 191–196.

Swets, D.L., Weng, J., 1996. Using discriminative eigenfeatures for image retrieval. IEEE Trans. Pattern Anal. Machine Intell. 18 (8), 831–836.

van't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.M.H., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., Schreiber, G.J., Kerkhoven, R.M., Roberts, C., Linsley, P.S., Bernards, R., Friend, S.H., 2002. Gene expression profiling predicts clinical outcome of breast cancer, letters to nature. Nature 415, 530–536. [Data Source: <http://www.rii.com/publications/2002/vantveer.html>].

Ye, J., 2005. Characterization of a family of algorithms for generalized discriminant analysis on undersampled problems. J. Mach. Learn. Res. 6, 483–502.

Yeoh, E.J., Ross, M.E., Shurtleff, S.A., Williams, W.K., Patel, D., Mahfouz, R., Behm, F.G., Raimondi, S.C., Relling, M.V., Patel, A., Cheng, C., Campana, D., Wilkins, D., Zhou, X., Li, J., Liu, H., Pui, C.H., Evans, W.E., Naeve, C., Wong, L., Downing, J.R., 2002. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. Cancer 1(2), 133–143. [Data Source: <http://www.stjuderesearch.org/data/ALL1/>].

Yu, H., Yang, J., 2001. A direct LDA algorithm for high-dimensional data-with application to face recognition. Pattern Recognition 34, 2067–2070.